Multiresolution alerting for motion detection*

Sylvia Gil, Thierry Pun

Computer Science Department
University of Geneva
24, rue Général Dufour
CH - 1211 Geneva 4
e-mail: gil@cui.unige.ch

Abstract

In this paper an alerting system is presented which aims at detecting moving objects in an image sequence. A preliminary estimate of moving objects is first obtained by temporal processing of the frames. This estimate is then iteratively refined by a multiresolution relaxation process until a precise binary mask is obtained, which selects the moving objects. Results for real image sequences are presented.

Résumé

Cet article présente un système d'alerte qui a pour but de détecter les objets en mouvement dans une séquence d'images. Une estimation préliminaire des objects en mouvement est d'abord obtenue au moyen d'opérateurs temporaux. Cette estimation est alors perfectionnée à l'aide d'une relaxation sur plusieurs résolutions, jusqu'à ce qu'un masque binaire qui sélectionne les objects en mouvement soit obtenu. Les résultats sont présentés pour des séquences d'images réelles.

1. Introduction

In this paper, an alerting system is proposed aiming at significantly reducing the amount of data where motion computation operators have to be applied. Its goal is to analyze an image sequence and to provide a set of spatio-temporal masks where motion actually occurs. In this way, motion evaluation can be restricted to these masks whose pixels are labelled as *dynamic*.

Various motion detection methods have been proposed based on the difference between subsequent frames of a sequence. For example, in [1] a method is presented to compute global thresholds in order to segment images into *static* and *dynamic* areas. Global thresholds are computed according to the noise probability density function (pdf). However, segmentation through global thresholding does not provide well segmented masks, due to false alarms and non-compact shapes. A local refinement is then applied on this preliminary data, based on the maximum-a-priori criterion (MAP). However, the use of global methods and the lack of a multiresolution signal representation lead to an oversegmentation, i.e. to many isolated masks which are actually part of the same moving object. In [2] another method based on MAP criterion is presented for the refinement of difference images. In this case an energy function is minimized using a stochastic iterative algorithm. The energy function takes into account spatial and temporal coherence. Stochastic relaxation requires intensive machine resources though.

The following section reports some details about the pyramidal representation. and temporal processing. The relaxation algorithm is presented in Section 3. Finally, Section4 presents some experimental results.

2. Temporal processing

For each image of the input sequence $I(x, y, t)$, an N-levels pyramid is constructed, from the highest resolution σ_0 (the image itself) at the bottom of the pyramid, to the coarsest resolution σ_{N-1} at its top. We use the notation $I(x, y, \sigma, t)$ with $\sigma = \sigma_0..\sigma_{N-1}$. These representations are computed by decomposing

*This work is supported by the Swiss National Fund for Scientific Research, Swiss National Research Program 23: Artitcial Intellignece and Robotics, grant No. 4023-27036.

the input image into a set of basis functions, in our case a set of β-*splines* functions [3].

Temporal differences have been widely used in the literature for motion detection (see for example [1][2]). In our system, temporal changes are evaluated for each pyramid level over ν frames, in order to have a preliminary estimate about the shape and the position of moving objects. Temporal differences are performed at each pixel $I(x, y, \sigma, t)$ along the temporal direction:

$$D_k(x, y, \sigma, t) = I(x, y, \sigma, t) - I(x, y, \sigma, t+k) \quad . \qquad (1)$$

From $D_k(x, y, \sigma)$ two types of contributions can be derived. The first one is computed according to the following expression:

$$C_1(x, y, \sigma) = \sum_{k=1}^{\nu} |D_k(x, y, \sigma)| \; . \qquad (2)$$

Theoretically this expression should be zero at locations where no motion occurs, and segmentation between static and dynamic regions should be straightforward. However, because of the presence of noise in image sequences, no credit may be given to weak values of contribution C_1 and only high values may be taken into account. In [1], the a-priori knowledge of the noise level is used to find the appropriate thresholds which segment *dynamic* from *static* regions. Therefore smooth edges (under the threshold) generated by moving objects will not be taken into account, making the recovery of compact shapes a difficult task.

In order to recover part of the weak edges which characterize slightly textured surfaces, a contribution $C_2(x, y, \sigma)$ has been introduced and is based on the sign changes in $D_k(x, y, \sigma)$. Figure 1 shows for the 1-D case how moving edges can be detected by their sign changes. Sign changes appear when edges are of the type *rect(x)*, therefore only this kind of edge can be recovered by this contribution. The task of C_2 is to find the locations in difference images $D(x, y, t, \sigma)$ where sign changes occur inside a sliding window (3x3 for all resolutions). When located, values generating the sign change have to be greater than a threshold τ in order to avoid noisy perturbations. τ is chosen according to the a-priori known level of noise in the sequence. Contributions C_1 and C_2 are finally summed into the motion estimate $E(x, y, \sigma)$. Figure 2 shows separately the two contributions and the estimate $E(x, y, \sigma)$. The first contribution shows spatio-temporal gradients, while the second one represents low edges characterizing the texture of moving objects

3. Pyramidal relaxation

Multi-grid or pyramidal relaxation methods have been successfully applied to a variety of vision problems [4][5]. A relaxation method with original incrementing functions is proposed in order to obtain compact and significant shapes of moving objects.

Pyramidal relaxation performs an integration of images belonging to different pyramid levels combining high resolution estimates (useful to constraint the precise shape of the objects) with low resolution ones (used to reduce the aperture problem by filling in the interior of the shapes). Since compact and meaningful shapes are desired, simultaneously horizontal and vertical relaxation processes are applied on the pyramid. The goal of the horizontal and vertical relaxation is to locally propagate the pixel values within each level as well as across the levels of the pyramid. In this way, spatial coherence is forced in intra- and inter-level neighborhoods of the pyramid. The horizontal relaxation is achieved by applying a

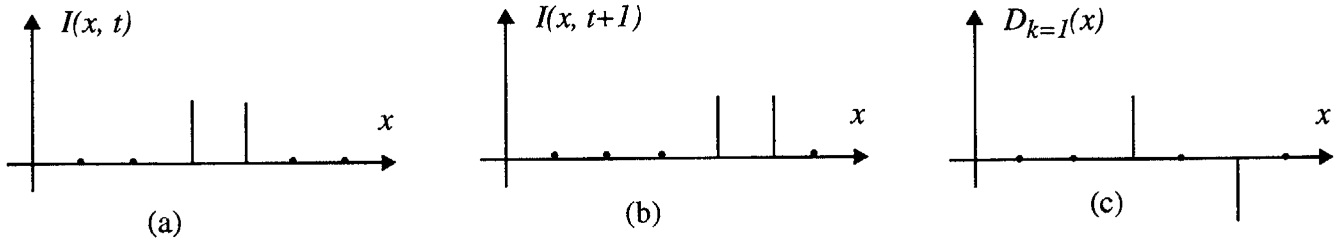

Figure 1: Characterization of moving edges of the type *rect(x)* by means of the sign change in the difference image: (a) moving edge of the type rect; (b) moving edge after 1 time unit; (c) sign change in the difference image $D_k(x)$.

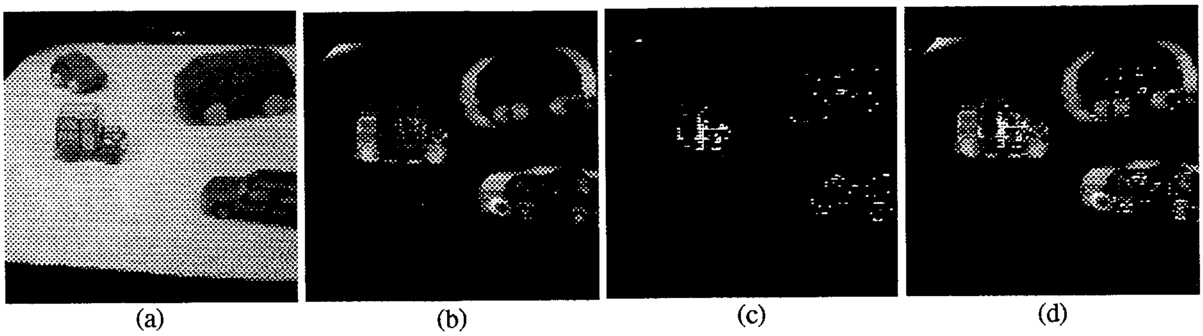

Figure 2: Results of temporal processing: (a) first frame of the image sequence; (b) contribution C_1; (c) contribution C_2 and (d) estimate of moving objects $E(x, y, \sigma)$.

Gaussian convolution operator on each pixel $E(x,y,\sigma)$ of the pyramid. Vertical relaxation is performed by propagating the pixel values from the top of the pyramid down to its bottom. This is done by repeatedly modifying the values of level σ_{k-1} according to level σ_k, until the bottom level σ_0 is reached. Each pixel $E(x,y,\sigma)$ of the σ_k-th level is used to update the values of four pixels of the lower image, i.e. $E(2x+i, 2y+j, \sigma_{k-1})$, $i,j \in \{0,1\}$. The updating consists of an additive increment Δ added to the values of each of these four pixels.

The increment Δ takes into account the values of the modifying pixel $E(x,y,\sigma_k)$ as well as the modified one $E(2x+i, 2y+j,\sigma_{k-1})$, and can be written as a product of two factors: $\Delta = \Delta_1 * \Delta_2$. To compute Δ_1, the range of values of the initial estimate $E(x,y,\sigma)$ is analyzed and split into two intervals by a threshold T, computed according to an a-priori known estimate of the image noise. The lower half interval (below T) defines *static* locations, whereas values in the higher one are associated with *dynamic* locations. The value of Δ_1 is defined by:

$$\Delta_1 = \begin{cases} -k_1 \cdot (E(x,y,\sigma_k) - T)^2 & \text{if } E(x,y,\sigma_k) \leq T \\ sigm(E(x,y,\sigma_k) - k_2 \cdot T) & \text{if } E(x,y,\sigma_k) > T \end{cases} \quad (3)$$

where k_1 and k_2 are positive constants used to bound $|\Delta_1|$ in [0,1] and to enforce continuity at T up to the first derivative; *sigm* is a sigmoid function of the type $sigm(x)=1/(1+c_1.\exp(-c_2.x))$. As a function of $E(x,y,\sigma_k)$, Δ_1 corresponds to a parabolic arc for values below T and to a sigmoid arc for values exceeding T.

In order to increment the values of $E(2x+i, 2y+j,\sigma_{k-1})$ without changing their dynamic range, a second factor Δ_2 is computed according to the following expression:

$$\Delta_2 = \begin{cases} E(2x+i, 2y+j, \sigma_{k-1}) & \text{if } \Delta_1 \leq 0 \\ M - E(2x+i, 2y+j, \sigma_{k-1}) & \text{otherwise,} \end{cases} \quad (4)$$

where $M = \max E(x,y,\sigma_{k-1})$. Therefore Δ_2 represents the maximum allowed increment. If Δ_1 represents a negative increment, Δ_2 corresponds to the maximum decrement, i.e. the value of $E(2x+i, 2y+j,\sigma_{k-1})$ itself.

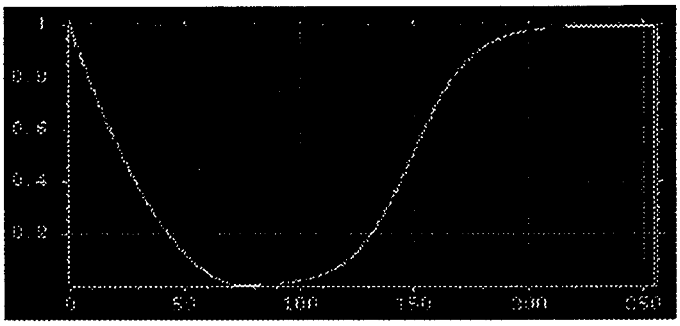

Figure 3: Increment $|\Delta_1|$ as a function of the grey level $E(x,y,\sigma_k)$. When its value is close to the uncertainty threshold T, $|\Delta_1|$ is close to zero, otherwise Δ_1 gets close to 1.

The relaxation process using $\Delta = \Delta_1 * \Delta_2$ is repeated until convergence is reached. A small number of iterations are necessary (1-3) to reach stability. Convergence is forced by decreasing the size of the Gaussian operator for the diffusion process. In this way, the values are allowed to change rapidly during the first iterations and then become more stable as the size of the spatial neighborhood is reduced.

4. Experimental results

Two real image sequences are used to show the results of the relaxation algorithm. Two frames were used to compute $E(x,y,\sigma)$, 4 resolution levels where used to build the pyramid and the full resolution images size is 256 x 256.

The first sequence shown in the first row of Figure 4 contains four toy-cars on a carpet, three of which are moving towards the center of the image. Also, the carpet illumination changes dramatically at the top-left corner of the image, creating an illusion of motion. Although objects are well contrasted the sequence is very noisy. The aperture problem appears clearly for the upper-right car since high motion estimates are very sparse. Results show that all moving objects have been detected and that their shape is compact.

The second sequence represents «Miss America» talking. She moves slightly her mouth and her right eye as she swings her head. The sequence presents a low noise level and the background is not very contrasted from the face. The estimate $E(x,y,\sigma)$ is very poor but, since the noise level is low this allows a low decision threshold. The resulting mask is not as compact as the previous sequence. This is due partly because the head does not move widely. Localization is accurate: her head, hair, shoulder and neck are segregated from the background.

Acknowledgments

We are grateful to R. Kumar and H.S. Sawhney at University of Massachusetts for providing the «cars» sequence. We would like to thank Michael Unser for his advice and his generous help. We also thank Harry Wechsler for his helpful comments and suggestions.

References

[1] T. Aach, A. Kaup, R. Mester, "Statistical model-based change detection in moving video", Signal Processing, Vol. 31, 1993, pp. 165-180.

[2] P. Lalande, P. Bouthemy, "A statistical approach to the detection and tracking of moving objects in an image sequence", Proc. of the Eusipco, Spain, 1990, pp. 947-950.

[3] M. Unser, A. Aldroubi, M. Eden, "The L2 polynomial spline pyramid", IEEE Trans. Patt. Anal. Mach. Intell., Vol. 15, No. 2, February 1993.

[4] D. Terzopoulos, "Image analysis using multigrid relaxation methods", IEEE Trans. Patt. Anal. Mach. Intell, Vol. 8, No. 2, March 1986.

[5] A. Rosenfeld, "Multiresolution image processing and analysis", Springer-Verlag, 1984.

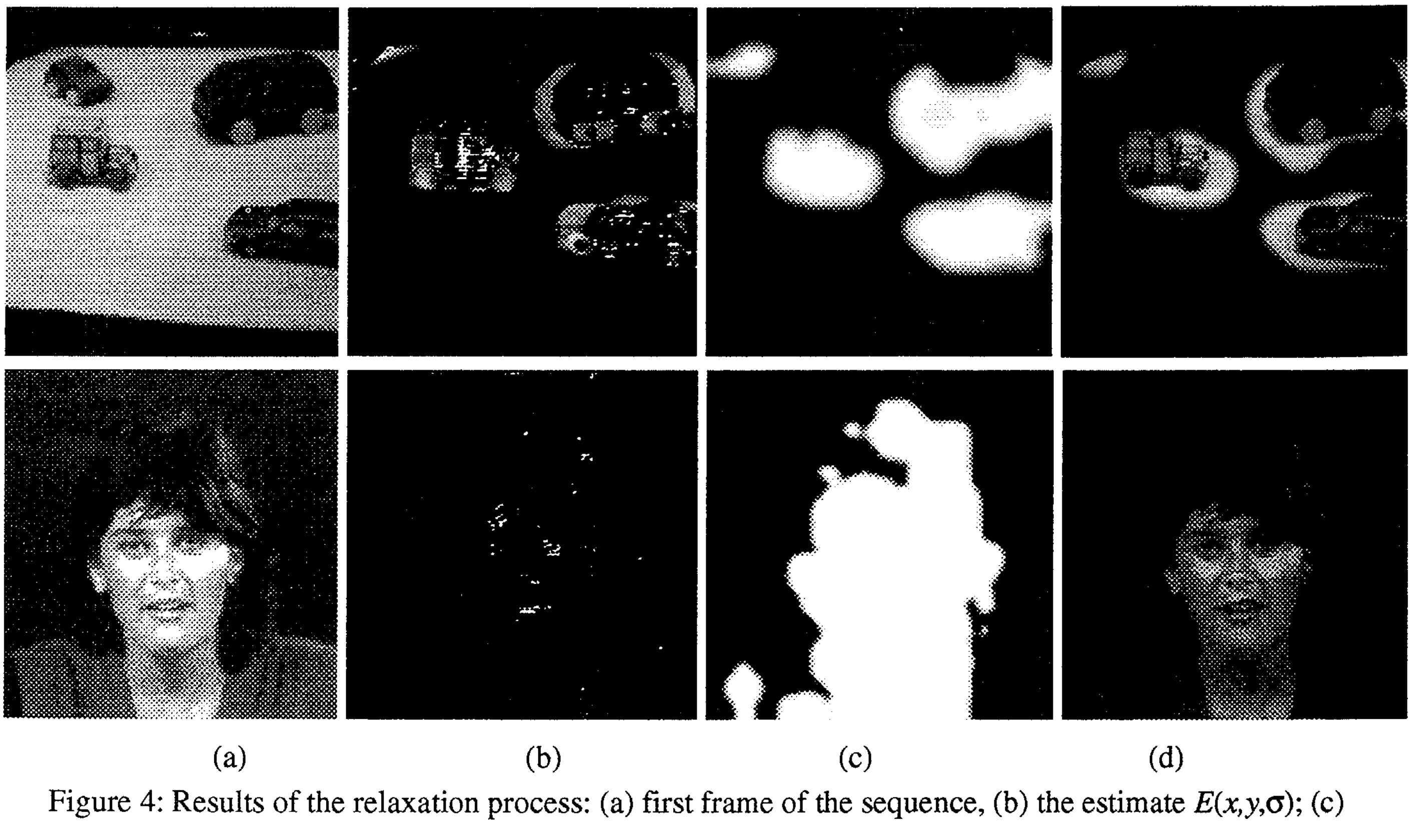

Figure 4: Results of the relaxation process: (a) first frame of the sequence, (b) the estimate $E(x,y,\sigma)$; (c) the binary masks; (d) the product between the mask and the first frame of the sequence (full resolution images).